Issues of nonlinearity and non-gaussianity

Contribution to the WWRP/THORPEX workshop on
4D-Var and ensemble Kalman filter inter-comparisons,
Buenos Aires, 10-13 November 2008

Marc Bocquet*

5 December 2008

From the point of view of mathematical modeling, a data assimilation system consists in a statistical description of a dynamical model (which can be stochastic or not), and of the related observations. The statistics on this system are meant to represent the uncertainty on the true state estimation. The proper statistical modeling depends on how uncertainty evolves under the full data assimilation system. Truncating statistics to the first and second order moments (mean and covariance matrix) leads to a (minimally committed) Gaussian modeling of this statistical system. This truncation is made necessary not only because of the complexity of the fully Bayesian data assimilation algorithms, but also because of the amount of information to be stored (filtering approach especially). However, this is also justified from the point of view of the evolution of the dynamical model. If, in the vicinity of a trajectory, the model can be replaced by its tangent linear, then initial Gaussian statistics will remain so in this vicinity. Non-linearities in the model and insufficient and infrequent observational data are likely to make the statistics diverge from ideal gaussianity.

The talks and the subsequent discussion of the session *Issues of nonlinearity and non-gaussianity* discussed these questions and the implementation of related ideas.

1 Why not non-Gaussian from the start ?

From an analytical point of view, regardless of the algorithmic complexity and numerical cost, the full statistical estimation problem can be solved by exact methods (smoothing or filtering problem) [10]. Monte-Carlo methods, such as the particle filters, have been built to offer a numerical solution of the full filtering equations.

These techniques are very efficient for small system sizes (less than 10 variables). But the numerical complexity has been shown to scale exponentially with the system size, or the innovation variance [12]. The most popular particle filter is called the *bootstrap filter* [4]. A major flaw of any particle filter is the degeneracy of the weights attached to each member of the ensemble. These weights measure the likelihood of a particle to comply

*CEREA joint laboratory École des Ponts ParisTech / EDF R&D, Université Paris-Est and INRIA. E-mail: bocquet@cerea.enpc.fr

with the observations. For high-dimensional systems, most of the weights are vanishing while one particle remains likely. Therefore the particle filter becomes useless for the estimation problem. In order to mitigate this effect, a resampling step is used though it does not fundamentally solve the issue.

To illustrate this point, a bootstrap filter is compared to an ensemble Kalman filter [EnKF] on a Lorenz 1996 [8] model with as few as 10 variables. Still, the particle filter requires 2×10^4 members to match the EnKF performance ! Results are shown on Fig.1.

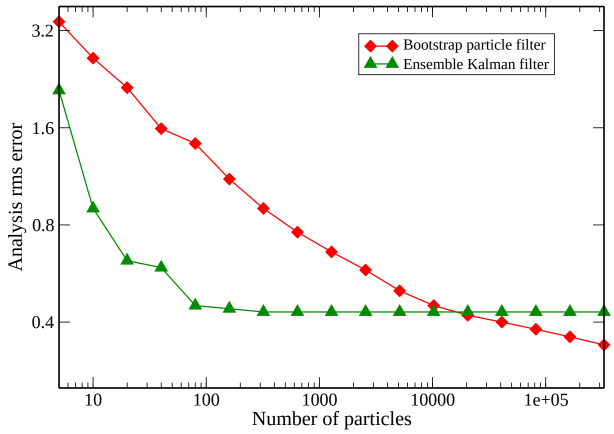

Figure 1: Comparison of the performance of a bootstrap filter with an ensemble Kalman filter on a Lorenz 1996 model with $N = 10$ variables, as the ensemble size is increased.

Therefore at least with basic (though not naive) algorithms, particle filters are still out of the question to take non-Gaussian effects of high-dimensional data assimilation system into account.

2 Dealing with non-gaussianity in a Gaussian framework

This explains the current way of dealing with nonlinearity and non-gaussianity: reasonable data assimilation should consider non-Gaussian effects as corrections to a Gaussian-based strategy. Variational approaches (4D-Var essentially) and ensemble-base Kalman filters have different constraints and take properly into account, though differently, the nonlinearities of models.

Non-gaussianity sources have been categorized into two families

- Nonlinearities in models which may come from: nonlinearity of Navier-Stokes equations leading to chaos, thresholds of microphysics (cloud, rain), chemistry of atmospheric compounds, increase in the model resolution (precipitation at convective scale), nonlinearity of observation operator model (satellite applications especially), etc. [1].

- Non-Gaussian priors which are sometimes more adequate description of the background: this is so for humidity (though a Gaussian anamorphosis is still possible), or for emission inventories in atmospheric chemistry. Observation error priors can also require non-Gaussian modeling: Huber norm which is a combination of l_1 and l_2 accounting for gross errors, l_∞ norm, log-normal and multiplicative errors, etc.

$\longrightarrow$ Read Mike Fisher and Shu-Chih Yang' contributions/presentations pdf for thorough illustrations of how nonlinearities and non-gaussianities can be dealt with in 4D-Var and ensemble-based Kalman filters.

3 Measuring non-gaussianity: how much do we loose being Gaussian ?

In a Gaussian framework, one needs a tool to assess the deviation from gaussianity mainly induced by nonlinearities of the model. One rigorous objective function to measure this deviation is the relative entropy [6]

$$\mathcal{K}(p, q) = \int \mathrm{d}p \ln \frac{p}{q} \,. \tag{1}$$

where p is the full probability density function [pdf] of the uncertainty on the system, and q is the Gaussian pdf that has the same first and second-order moments. It is hoped that p can be assessed by the use of an ensemble, such as the one used by ensemble-based filters. This measure has been advocated in the field of predictability by [5]. It is however not so easy to perform such estimation for high-dimensional systems, especially with a small ensemble. Several solutions have been proposed: compute relative entropies of marginals of p, or compute expansions of relative entropy. The latter expansions are based on Gram-Charlier or Edgeworth expansion of p/q. These expansions depend on skewness and kurtosis, which is consistent with the use of skewness, kurtosis or both in the diagnosis of non-normality.

Other tests of null-hypothesis can be used, such as univariate tests (Kolmogorov-Smirnov, Anderson-Darling, Shapiro-Wilk, etc.). To exploit them in high dimensions, a mapping must be defined between multivariate system states to a scalar, using for instance a Mahalanobis norm leading to a χ^2 law.

The loss of gaussianity of an ensemble in the flow of Lorenz 1963 model [7] is illustrated on Fig.2. Gaussianity is no longer an adequate description beyond $t = 0.5$.

4 Bridging the gap between Gaussian and non-Gaussian data assimilation

There have been recent attempts to make use of non-Gaussian ideas in geophysical (or geophysically-inspired) data assimilation. They remain quite specific in their application, because of their underlying hypotheses. They are nevertheless promising. We would like to mention a few.

Statistical expansion about the climatology Considering an ensemble-based filtering, one idea is to still perform perform a statistical analysis of the ensemble with the mean and error covariance matrix. But instead of forming a Gaussian pdf as a prior for the analysis, one computes the *closest* probability density function to the climatology with

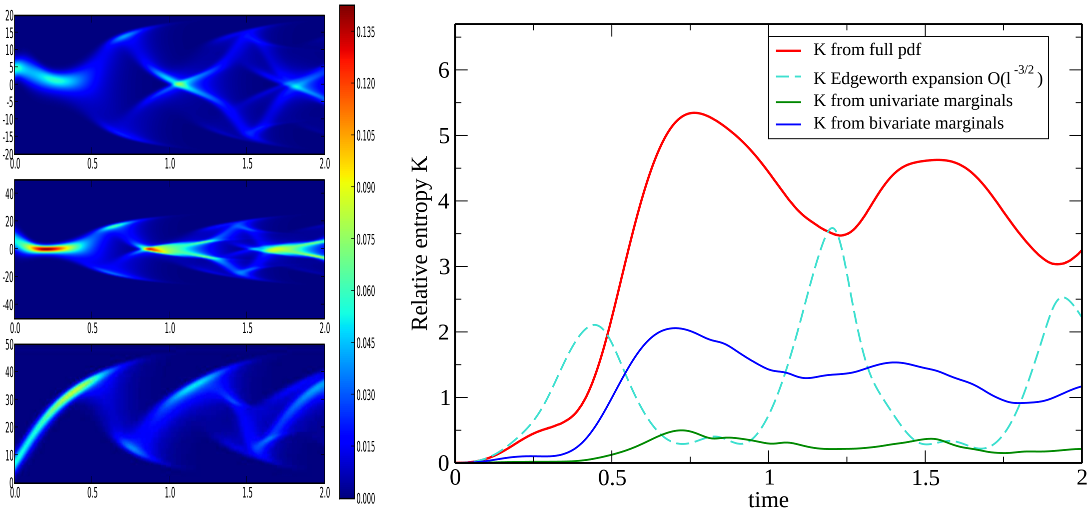

Figure 2: On the left: the density of an initially Gaussian ensemble in the flow projected on the three variables of the model. On the right is plotted the relative entropy of the same ensemble statistics, the mean of the univariate marginals entropies, and an expansion of the full pdf relative entropy (skewness order only).

the same first and second-order moments [3]. The measure of distance between a pdf and a climatology is provided by the relative entropy.

Using non-Gaussian deviations in the priors to improve analysis Given the statistics of innovations, it is possible to compute the mean and covariance matrix but also deviation from gaussianity (skewness and kurtosis). Given those higher order statistics on the innovations, Pires and Talagrand have proposed to estimate a joint prior on the observation and background errors, using the maximum entropy principle. Beyond gaussianity, these priors incorporate higher order corrections. This approach has been shown to be beneficial to the subsequent analysis.

Particle filtering with Ensemble Kalman Filter as an importance proposal From the point of view of particle filtering, the full pdf of a data assimilation system is approximated by the P members ensemble discrete pdf:

$$p_t(\boldsymbol{X}_t|\boldsymbol{Y}_t) \simeq \sum_{i=1}^{P} w_t^i \delta(\boldsymbol{X}_t - \boldsymbol{X}_t^i)\,, \tag{2}$$

where $\boldsymbol{X}_t$ is a complete state trajectory up to time t, $\boldsymbol{Y}_t$ is the complete observations set up to time t, and w_t^i are the weights attached to each particle $\boldsymbol{X}_t^i$. There is a degree of freedom in this representation: the weights and the collection of particles, which can be exploited. In the Monte-Carlo literature, this is known as importance sample. Indeed the particles can be drawn from (essentially) an arbitrary pdf, provided the weights are corrected so that the discrete pdf still be representative of the system uncertainty. The importance proposal pdf of the *bootstrap filter* is the transition kernel of the model. In [9], a particle filter has been built with an importance sampling given by a Kalman filter which tracks the system. Contrary to the bootstrap filter, only particles with a reasonable likelihood given the observations will be sampled. (This type of approach has also been advocated by P.J. Van Leeuwen.) Using an ensemble Kalman filter to do so, as suggested in [11], one can define a true particle filter with weights attached to each member of the ensemble. If the implementation is similar to EnKF, the asymptotics are those of the non-Gaussian Bayesian filtering equations (therefore fully non-Gaussian). In Fig.3, a test is performed to compare the merits of an EnKF, the bootstrap filter and the weighted Kalman filter [WenKF]. If WenKF still requires a large number of particles and resampling, the progress is spectacular as compared to the bootstrap filter.

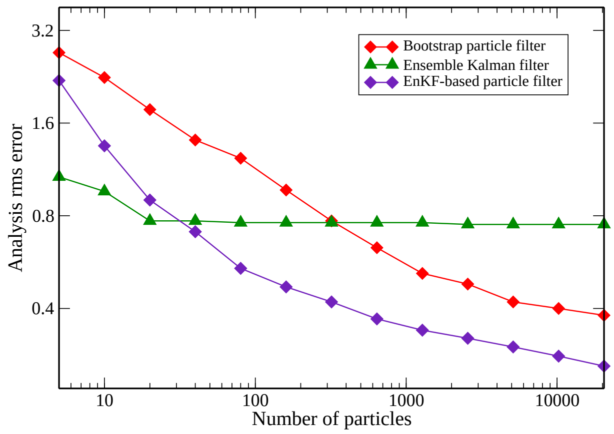

Figure 3: Comparison of the performance of three filters (EnKF, bootstrap and WenKF) on a Lorenz 1996 model with $N = 5$ variables.

Fully non-Gaussian methods for high-dimensional linear models They are geophysical relevant cases where the models are approximatively linear. But the prior may not be naturally modeled as Gaussian. This is the case for tracer dispersion. In that situation (non-Gaussian priors, linear models), a non-Gaussian analysis can be performed without approximation, using nonlinear convex analysis. It usually yields non-quadratic cost-functions that generalize 4D-Var and PSAS in the specific linear model case [2]. An example is given on Fig.4, where data assimilation on real tracer concentration measurements is performed with Gaussian and Non-Gaussian approaches. Besides better statistical scores, the contour of the plume is much neater and closer to the reference in the non-Gaussian method case.

5 Conclusions

Fully non-Gaussian numerical solutions of the estimation problem, such as particle filters, are *still* not affordable, though significant improvement cannot be ruled out in the near future [13], in the same vein as the weighted ensemble Kalman filter tested here. Sticking to Gaussian assumption based data assimilation methods, mathematical tools exist that can *objectively* measure the departure from gaussianity, and hence give a proper estimation

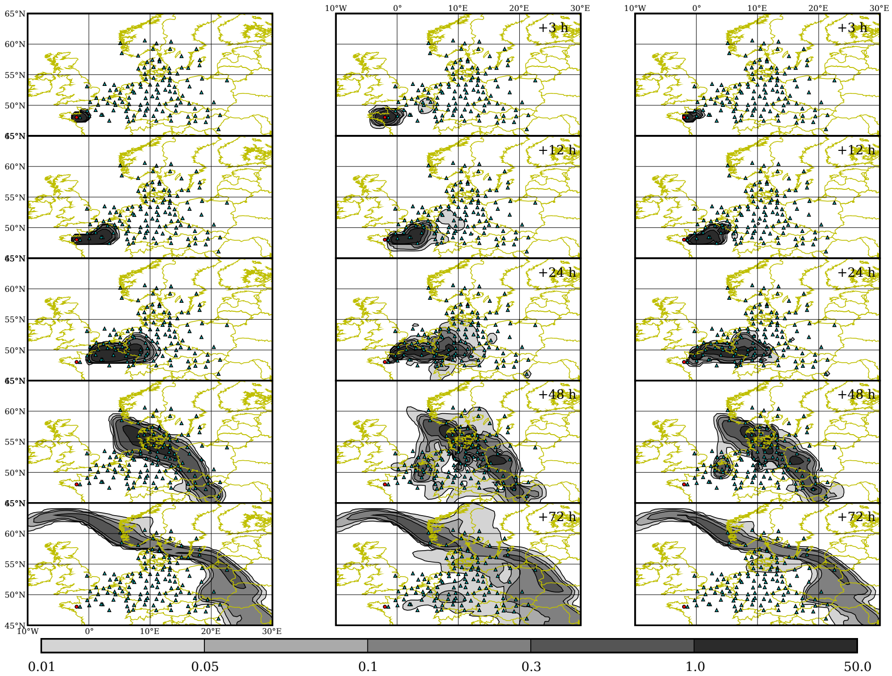

Figure 4: On the left: The reference tracer dispersion event ETEX-I simulated knowing the source. Middle: Gaussian-based 4D-Var with sequentially assimilated true observations. Right: non-Gaussian assimilation approach. Units are ng.m^{-3}.

of the reliability of the Gaussian assumptions. New methods are now proposed in specific cases for the time-being, and sometimes even in high dimensions, which can be considered as*non-perturbative* and handle non-gaussianity (maximum entropy filter, maximum entropy inference, non-Gaussian priors constructions).

However with increasing resolution, uncertainty might soon be considered locally Gaussian to a very good approximation. So do we need (or will we need) non-Gaussian modeling after all ? If so, non-Gaussian approaches will remain just refinements of the mainstream Gaussian methods because only small deviations from gaussianity have to be considered. Nevertheless, this idea do not extend to the modeling of truly non-Gaussian priors (positive variables, multi-modal pdf).

The concepts presented here were orientated towards getting the best estimator (mean of analysis forecast). The point of view changes when truly facing the estimation of the pdf (*best ensemble estimate*), or at least higher moments, which is the natural view point of ensemble prediction. Then non-Gaussian modeling may become a very significant issue since 4D-Var and Ensemble-based Kalman filtering are by construction insufficient for high-order moments corrections.

References

[1] E. Andersson, M. Fisher, Hòlm E., L. Isaken, G. Radnóti, and Y. Trémolet. Will the 4D-Var approach be defeated by nonlinearity ? Technical Report 479, ECMWF,

September 2005.

[2] M. Bocquet. Inverse modelling of atmospheric tracers: Non-Gaussian methods and second-order sensitivity analysis. *Nonlin. Processes in Geophys.*, 15:127–143, 2008.

[3] G. L. Eyink and S. Kim. A maximum entropy method for particle filtering. *J.Stat.Phys*, 123:1071–1128, 2006.

[4] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE-F*, 140:107–113, 1993.

[5] R. Kleeman. Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.*, 59:2057–2072, 2002.

[6] S. Kullback. *Information theory and statistics.* New York: Wiley, 1959.

[7] E.N. Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20:130–141, 1963.

[8] E.N. Lorenz and K.E. Emmanuel. Optimal sites for supplementary weather observations: simulation with a small model. *J. Atmos. Sci.*, 55:399–414, 1998.

[9] R. Merwe, A. Doucet, N. Freitas, and E. Wan. The unscented particle filter. Technical Report Report CUED/F-INFENG/TR 380, Cambridge University Engineering Department, August 2000.

[10] R.N. Miller, E.F. Carter, and S.T. Blue. Data assimilation into nonlinear stochastic models. *Tellus A*, 51:167–194, 1999.

[11] N. Papadakis. *Assimilation de données images: application au suivi de courbes et de champs de vecteurs.* PhD thesis, Université de Rennes I, 2007.

[12] C. Snyder, T. Bengtsson, P. Bickel, and J. L. Anderson. Towards a nonlinear ensemble filter for high-dimensional systems. *Mon. Wea. Rev.*, 2008.

[13] P. J. Van Leeuwen. A variance-minimizing filter for large-scale applications. *Mon. Wea. Rev.*, 131:2071–2084, 2003.